DISTRIBUTION SHEET

To CDMS Team	From W. T. Rohen <i>W.T. Rohen</i>	Page 1 of 1
Project Title/Work Order HIMACS File Management Software and Strategies		Date 09/30/94
		EDT No. 604043
		ECN No.

Name	MSIN	Text With All Attach.	Text Only	Attach./ Appendix Only	EDT/ECN Only
------	------	-----------------------------	-----------	------------------------------	-----------------

E. R. Carroll	B5-26	X			
R. E. Harding	A2-35	X			
M. M. Ledgerwood	A2-90	X			
B. D. Newman	A2-35	X			
J. O. Perkins	A2-90	X			
W. M. Perrine	B5-25	X			
J. R. Phillips	A2-90	X			
J. J. Ratcliffe	B5-25	X			
W. T. Rohen	B5-26	X			
R. M. Yeats	H5-14	X			

OSTI (2)
Central Files (2 + Copy)
18-07 x(2)
18-04 2

DISCLAIMER

Portions of this document may be illegible in electronic image products. Images are produced from the best available original document.

Sta #10 NOV 16 1994	Page 1 of 1 1. EDT 604043
-------------------------------	-------------------------------------

ENGINEERING DATA TRANSMITTAL

2. To: (Receiving Organization) Distribution List	3. From: (Originating Organization) ISS/HIMACS System Development	4. Related EDT No.:
5. Proj./Prog./Dept./Div.: HIMACS	6. Cog. Engr.: W. T. Rohen	7. Purchase Order No.:
8. Originator Remarks: This document contains the File Management Strategies for the Hanford Information and Control Strategies Program.		9. Equip./Component No.:
11. Receiver Remarks:		10. System/Bldg./Facility:
		12. Major Assm. Dwg. No.:
		13. Permit/Permit Application No.:
		14. Required Response Date:

15. DATA TRANSMITTED					(F)	(G)	(H)	(I)
(A) Item No.	(B) Document/Drawing No.	(C) Sheet No.	(D) Rev. No.	(E) Title or Description of Data Transmitted	Approval Designator	Reason for Transmittal	Originator Disposition	Receiver Disposition
1	WHC-SD-GN-SP-30001		0	HIMACS File Management Software and Strategies	N/A	1	1	1

16. KEY											
Approval Designator (F)		Reason for Transmittal (G)				Disposition (H) & (I)					
E, S, Q, D or N/A (see WHC-CM-3-5, Sec.12.7)		1. Approval 4. Review 2. Release 5. Post-Review 3. Information 6. Dist. (Receipt Acknow. Required)				1. Approved 4. Reviewed no/comment 2. Approved w/comment 5. Reviewed w/comment 3. Disapproved w/comment 6. Receipt acknowledged					
(G)	(H)	17. SIGNATURE/DISTRIBUTION (See Approval Designator for required signatures)						(G)	(H)		
Reason	Disp.	(J) Name	(K) Signature	(L) Date	(M) MSIN	(J) Name	(K) Signature	(L) Date	(M) MSIN	Reason	Disp.
1		Cog.Eng./W. T. Rohen		10-25-94	BS-26						
1		Cog. Mgr./E. R. Carroll		10/25/94	BS-26						
1		QA									
		Safety									
		Env.									
1		Sponsor									

18. W. T. Rohen Signature of EDT Originator Date 10-29-94	19. E. R. Carroll Authorized Representative Date 10/25/94 for Receiving Organization	20. E. R. Carroll Cognizant Manager Date 10/25/94	21. DOE APPROVAL (if required) Ctrl. No. ☐ Approved ☐ Approved w/comments ☐ Disapproved w/comments
--	---	---	---

RELEASE AUTHORIZATION

Document Number: WHC-SD-GN-SP-30001, Rev. 0

Document Title: HIMACS Files Management Software and Strategies

Release Date: 11/3/94

**This document was reviewed following the
procedures described in WHC-CM-3-4 and is:**

APPROVED FOR PUBLIC RELEASE

WHC Information Release Administration Specialist:

Chris Willingham

C. Willingham

11/3/94

TRADEMARK DISCLAIMER. Reference herein to any specific commercial product, process, or service by trade name, trademark, manufacturer, or otherwise, does not necessarily constitute or imply its endorsement, recommendation, or favoring by the United States Government or any agency thereof or its contractors or subcontractors.

This report has been reproduced from the best available copy. Available in paper copy and microfiche. Printed in the United States of America. Available to the U.S. Department of Energy and its contractors from:

U.S. Department of Energy
Office of Scientific and Technical Information (OSTI)
P.O. Box 62
Oak Ridge, TN 37831
Telephone: (615) 576-8401

Available to the public from:

U.S. Department of Commerce
National Technical Information Service (NTIS)
5285 Port Royal Road
Springfield, VA 22161
Telephone: (703) 487-4650

SUPPORTING DOCUMENT

1. Total Pages 19

2. Title

HIMACS File Management Software and Strategies

3. Number

WHC-SD-GN-SP-30001

4. Rev No.

0

5. Key Words

File, File Management, File Server
Tools
Process for Development
Repositories
File Naming Standards
Server Partitioning

6. Author

Name: W. T. Rohen

Signature

Organization/Charge Code 62420/J6240

7. Abstract

This document discusses the various file management technologies available for potential use with microcomputers at Hanford.

8.

RELEASE STAMP

TITLE: HIMACS FILE MANAGEMENT
SOFTWARE AND STRATEGIES

10/25/94
E. R. Carroll, Manager Date

WESTINGHOUSE HANFORD COMPANY

HIMACS

Hanford Information Management and Control Strategy

File Management
Software and Strategies

WHC-SD-GN-SP-30001

PREPARED BY

Boeing Computer Services Richland
ISS/HIMACS Systems Development
William T. Rohen

DISCLAIMER

This report was prepared as an account of work sponsored by an agency of the United States Government. Neither the United States Government nor any agency thereof, nor any of their employees, makes any warranty, express or implied, or assumes any legal liability or responsibility for the accuracy, completeness, or usefulness of any information, apparatus, product, or process disclosed, or represents that its use would not infringe privately owned rights. Reference herein to any specific commercial product, process, or service by trade name, trademark, manufacturer, or otherwise does not necessarily constitute or imply its endorsement, recommendation, or favoring by the United States Government or any agency thereof. The views and opinions of authors expressed herein do not necessarily state or reflect those of the United States Government or any agency thereof.

MASTER

DISTRIBUTION OF THIS DOCUMENT IS UNLIMITED

TABLE OF CONTENTS

1.0	INTRODUCTION	4
1.1	Purpose	4
1.2	Background	4
1.3	Goals of Architecture Development	4
1.4	Approach	4
1.5	Scope	4
2.0	SYSTEMS DEFINITION	5
2.1	Definition of Terms	5
2.2	Enabling Technologies	6
2.3	Benefits	6
2.4	Standards	6
3.0	AVAILABLE TOOLS	8
3.1	Purpose of Software Tools	8
3.1.1	Server Partitioning	8
3.1.2	Saros Mezzanine	8
3.1.3	Interleaf	9
3.1.4	File Viewers	10
3.1.5	Visual Basic/Paradox	10
3.1.6	Windows File Manager	10
3.1.7	Lotus Notes	10
3.1.8	Viewstar	11
3.1.9	Future Trends	11
3.1.10	Industry News	12
3.3	File Management and Information Repositories	12
3.4	Networking of Functions	13
4.0	PROCESS FOR DEVELOPMENT	14
4.1	Responsibilities	14
4.2	Funding Identification	14
4.3	Computer Program Interfaces	14
4.4	On-line Help	14
4.5	Training	15
4.6	Defining Automatic Metrics Collection	15
4.7	Technical and User Support	15
4.8	Administrative Support	15
APPENDIX A:	File Naming Standards	16
APPENDIX B:	Server Partitioning	18

1.0 INTRODUCTION

1.1 Purpose

The purpose of this document is to discuss various file management technologies available for potential use with microcomputers at Hanford.

1.2 Background

In support of Hanford Information Management And Control Strategy (HIMACS), a variety of engineers at Boeing Computer Services Richland (BCSR) explored new and existing technologies for managing files in a microcomputer environment. Technologies included:

- Network server shared areas, controlling read and read/write capabilities
- Saros Mezzanine
- Interleaf
- File viewers
- Visual Basic/Paradox
- Windows file manager
- Lotus Notes
- Viewstar
- Future Trends

1.3 Goals of Architecture Development

The goals were:

- Increase knowledge and skills of the technologies, complexities, and issues involved with file management in Personal Computer (PC) environments
- Determine potential needs and applications for the technologies
- Develop demonstratable examples of the technologies
- Discover and pursue issues related to the technologies
- Locate potential pilot projects involving the technologies
- Develop preliminary/proposed standards.

1.4 Approach

Due to limited resources (funds, training, labor, and equipment), trade journals and other projects were used, where possible, to supplement "test bed" evaluation of tools. Pilot projects were sought and worked on an as-available basis. Literature on the various products mentioned in this document are available via the author.

1.5 Scope

The File Management Technology studies were limited in nature, and are meant to be used as a foundation for follow-on projects. The document is high-level, with many subjective comments. The focus is on Disk Operating System (DOS) based PC's, using Windows software. While Macintosh, Unix, and X-Windows based workstations are in use at Hanford, the predominant platform is 386SX (or better) PC with DOS and Windows 3.1.

2.0 SYSTEMS DEFINITION

2.1 Definition of Terms

API	Application Program Interface. An API provides a standardized method of vertical communications between a protocol and an application program. In terms of the OSI 7-layer reference model, an API typically resides between the session and transport layers. Examples of APIs include NetBIOS, Named Pipes, and Sockets and Streams.
ASCII	American Standard Code for Information Interchange. The code corresponding to computer byte information. It is a 7-bit character encoding standard or protocol for displaying characters and transferring data between computers and associated equipment. ASCII codes are numbers from 0 to 255. The first 32 numbers are non-printing control codes (such as line feed, carriage return, and bell); the numbers 33 to 127 are for letters and punctuation; and the remaining 128 numbers called extended characters, can vary. They are usually used for accented letters, graphic, and other special characters.
BCSR	BCS Richland, Incorporated, a subsidiary of BCS (Boeing Computer Services).
BSEARCH	Business Services Electronic ARCHive. A BCSR document management system that uses Saros Mezzanine.
Compound documents	Documents containing multiple types of information and/or files.
DDE	Dynamic Data Exchange. A Windows messaging protocol and computer technology that allows data to be revised by one software product, and have changes dynamically appear in other products referencing the data.
DDE/OLE	See DDE and OLE. These are Microsoft technologies that allow different types of files (text, images, sound, graphics, etc.) to be linked together into compound documents. They enable files to be embedded or linked to other files, creating compound documents. The files can be seen as icons, launching the originating application when desired. For software to function properly, it must be capable of running DDE and OLE version 2.*, preferably in both client and server modes.
DOS	Disk Operating System. The prevalent base computer operating system for Intel-based computers.
ETEXT	The ESQD (Westinghouse department) Text Evaluation and eXchange Tool standalone prototype system. This system used TOPIC to demonstrate state-of-the-art capabilities and potential for a text search and retrieval system. It was superseded by SHARE.
GUI	Graphical User Interface. A way for humans to communicate with a computer that typically uses graphics mode instead of character mode. Usually involves a mouse. Alternative communication methods include command line interface and menu driven.
HIMACS	The Hanford Information Management And Control Strategy plan for the nineties. The plan involves the integration of network and work station operating systems, and application software into a controlled, unified system.
HLAN	Hanford Local Area Network. This is the term commonly referred to by PC users at Hanford. Technically speaking, it is not a local area network, therefore HAN may be a more appropriate term. It is a Metropolitan Area Network (MAN) that provides PC communications across the Hanford site.
HSD	BCSR's HIMACS Systems Development department, managed by Ed Carroll.
HUH	HLAN User's Help-n-hints. A Hanford Windows-based help file system .
Hypertext	A term that refers to linking a textual position in a document to a different textual position (either in the same or a different document). This is often used to link referenced documents, and to link table of contents to associated sections.
Infobase	A database comprised of textual documents, associated word indexes, and software that provides text search and retrieval of the documents.
ISEARCH	Hanford's Information Services Electronic ARCHival computer system.
ISS	Information and Scientific Systems. An organization of BCSR.
LAN	Local Area Network. A segment or interconnection of segments bounded by an active network device operating at or above the data link layer (e.g., minimally a bridging device). Note: A LAN also refers in literature to a geographical entity. Generally this implies a localized network, and perhaps a building. It is a localized communications network to provide computers with the capability to share and transfer data with other computers connected to the LAN.
mb	Abbreviation for megabyte. One million bytes, or one thousand kilobytes of computer storage.
Microsoft Windows	A method for providing users with a standard GUI for all programs. Once a user learns how to use Windows, all Windows-based applications function in a like manner, reducing the learning curve.
OCR	(1) Optical Character Reader. Hardware that performs the actual scanning and conversion. (2) Optical Character Recognition. Computer technology that reads and converts images into ASCII characters.
OLE	Object Linking and Embedding (or Extensions). A computer technology that allows software referencing data to automatically launch the software product that originally created the data. See Embedded Object and Linked Object.
PACE	Hanford's Procedure Access Controlled Electronically software system.

PC	Personal Computer. Typically an Intel-based computer.
PNL	Pacific Northwest Laboratories company, Battelle Memorial Institute, a Department Of Energy/Richland (DOE-RL) contractor involved in implementing a text search and retrieval system. Also known as Battelle.
PPC	Program/Project Planning and Control
RTF	Rich Text Format. Rich text means that the text is not limited to just ASCII characters. Instead, other information, such as pictures, DDE/OLE objects, etc. can be placed into the text. This is a Microsoft standard.
SARIS	PNL's Safeguards Active Response Inventory System
SGML	Standard Generalized Markup Language. A methodology to add structure to textual information.
SHARE	Search Hanford Accessible Reports Electronically, the production-implemented text search and retrieval system for ESQD. It uses TOPIC software, and was based on the ETEXT prototype.
SQL	Structured Query Language. This refers to the industry-standard relational database design, with standard database commands that can be commonly accessed by a variety of software products. It can provide a method for accessing structured data information. The structured information may also include the information pertaining to files (such as author, creation date, version, etc.)
TCP/IP	Transmission Control Protocol / Internet Protocol. A network software protocol will enable access to servers in a client/server mode, which will reduce network traffic and often improve program performance.
Text Retrieval	A feature that provides for retrieval of documents containing desired text, without the need for first creating keyword lists.
Unix	Not an acronym at all. Was a joke on MIT's MULTICS. It is a multi-user Operating System (OS) typically found on DEC and SUN computers, featuring multiprocessing capabilities. It was originally developed by AT&T in 1969.
VAR	Value Added Reseller
XNS	Xerox Network/National Standard/Service/Systems. XNS is a proprietary networking protocol, corresponding to the Open Systems Interconnection (OSI) network and transport layers.

2.2 Enabling Technologies

- Microsoft Windows
- DDE/OLE
- SQL
- TCP/IP

Please refer to the definitions in section 2.1, for an explanation of the above terminology.

2.3 Benefits

File management has been done in an unstructured manner. By adding structure, the amount of redundancy can be reduced, lost or over-written data can be minimized, previous versions can be maintained and retrieved, distribution of information can be centralized and controlled, and management can be standardized.

2.4 Standards

Standards that are in use at Hanford include SQL (Oracle and Sybase), WordPerfect 5.* for documents, cc:Mail for electronic mail, and XNS for network protocol.

Standards needed for file management include TCP/IP for client/server, upgraded cc:Mail (to allow DDE/OLE and shared name/address books), WordPerfect 6.* for documents (for OLE/DDE), SGML (both rules, and viewers and editors), and naming conventions (see appendix A).

A standard must also be established to acknowledge that electronically stored documents can be viewed as "original" and "record copy". Also, "electronic

authorization/approval" must become an accepted alternative to handwritten signature approval processes.

3.0 AVAILABLE TOOLS

3.1 Purpose of Software Tools

3.1.1 Server Partitioning

Server partitioning is a technology that has been in use at Hanford for over 10 years. It is a technique of setting up access to different shared areas, each pointing to separate file directories on a file server. The access can be passworded, and be set as either read-only or read-write.

The capability to partition servers could be used to create an electronic library. Each group would manage their own files (in read/write mode), while all could have read-only access to the whole. See Appendix B for an example of a proposed scheme.

The whole collection could then be indexed using a high-end software product, such as TOPIC, while the group areas could have group indexes created as necessary, using such small-end index products as WordPerfect's QuickIndex (included free as part of the version 6.* word processing software).

3.1.2 Saros Mezzanine

Saros Mezzanine is currently in use at Hanford for the BSEARCH project. Saros uses SQL as a backend for a full librarian system. User access control to documents and groups of documents is provided. Full text searching, and a document printer/viewer is included, and application launching to WordPerfect (and others) is also included. Conversely, while in WordPerfect, a user can be prompted to check-out documents from Saros.

Documents are renamed to encryptic names, and access to them is prevented except via Saros. The SQL database is capable of storing document names of 120+ characters, which is much greater than the 8.3 DOS naming convention.

As part of the file management evaluation, an evaluation copy of Saros was procured. A test collection of documents was created, consisting of tank farm records (supplied via Ray Moshofsky, 376-9197), Hanford Info, the HSD server area, and various other document sources. The test collection totaled 5,246 files (and 270 megabytes). This allowed for a realistic concept of performance issues.

During the creation of the system, several different methods were used to check in documents. These were Windows File Manager (to drag/drop), a bulk add script (basically an ASCII text listing of files), and Saros-created macros (for WordPerfect 5.2/Windows). All three methods worked, each served specific purposes. The main difficulty encountered was in adding more than 700 documents at a time. Since the words were being indexed at the same time that they were added to the library, the server was being multi-tasked to the point where the PC would "time out" for awhile. Other than that, and a few tuning tweaks, the Saros product seemed to work well.

The Saros products evaluated were Document Manager 1.0 (the user front-end) and FileShare 3.1 (administration utilities). Users can access the documents from a variety of work station platforms, including both Microsoft Windows and

Macintosh. The product is changing and growing rapidly. It is being incorporated into quite a few other products. The text retrieval engine currently included with Saros is Fulcrum. The TOPIC product is expected to be optional in a future version of Saros.

3.1.3 Interleaf

Interleaf was evaluated via product literature. Interleaf has been to Hanford several times, to provide demonstrations, and is available for additional demonstrations. Interleaf is viewed as an enterprise-wide product, in comparison to the Saros product (a work group solution). Interleaf is willing to provide their product at no cost for evaluation purposes.

Interleaf has three main products. "RDM" is their document management product, and costs \$50,000 for 16 users. Interleaf is their authoring tool, and costs \$2,500 for each authoring station. The authors convert documents into Interleaf, which also store them in compressed format. The third product is WorldView/Press, which is the viewer. It is \$5,000 for the DOS server product, and \$75 per user. All three can be purchased as the "Intellect" suite. The Intellect suite costs approximately \$130,000 to \$165,000 for 32 users, and includes 12 weeks of consultation and installation. The product is also available from Value-added resellers (VARs), so sole-source justification will be unnecessary.

RDM uses an Oracle engine, and has gone through six major revisions. Interleaf uses the Fulcrum text retrieval engine (see note number 3). It integrates SGML structure with legacy documents, producing Interleaf-formatted compressed information. There is a parser to aid in converting to SGML. There is also an integrated workflow product. Old versions can be maintained by storing as separate documents (different names). SGML implies automatic hyperlinking, and thus auto-rebuilding of hyperlinks as documents change. Print capability can be restricted and controlled, with an audit trail for document update/notification purposes.

- Fulcrum has several product levels. A basic level consists of simple word and boolean search. The "full set" level has additional features. I will soon be contacted by a Fulcrum representative with more details. Most imbedders of Fulcrum's API have chosen to incorporate some, but not all of the "full set" features. Saros will supply the "full set" for an additional fee of approximately \$40 per user.

Kurt Conrad (of BCSR, 376-9404) and Andy Alcorn (of PNL, 372-4794) have both been in contact with Interleaf. Kurt is very interested in Interleaf because of its SGML capabilities. Andy is a former Interleaf employee, and currently has copies of the Interleaf products for PNL evaluation and installation. Additionally, Siemens Power Corporation (of Richland, 375-8100) and Boeing Seattle (206-657-5902) use their products. Boeing is using the product for the PSDS system (which Ray Moshofsky of BCSR used to work on).

Boeing substituted "Online" text retrieval for Fulcrum. This demonstrates that other API-available text retrieval engines can also be substituted. WorldView 3 will be designed to be search engine independent (to accommodate OpenText, Online, and Topic). Boeing estimates cost savings for PSDS to be approximately

\$1.2 billion over five years. A Boeing-produced video is available from Bill Rohen, which highlights the PSDS system and Interleaf.

3.1.4 File Viewers

Windows can associate files with their originating applications, based on file extensions. This was discussed in section 2.4. However, the originating application typically allows modification of the information. This isn't always desired, especially for the dissemination of official, approved documentation where the company wants to prevent temporary modification prior to printing. In these instances, a file viewer/printer is necessary.

File viewers and printers can emulate the originating application enough to understand the proper formats for viewing and printing. Typically, all editing features have been removed. This provides a generic, relatively fast software program that is capable of viewing a variety of file formats.

Software products researched include WordPerfect's Envoy. Envoy is capable of viewing and printing a variety of file formats, without providing the users edit capability. This is valuable for controlling approved manuals. File formats include all WordPerfect versions, Microsoft Word, and ASCII text. Files are stored in compressed format, and can include the viewer as part of the file. The product includes hypertext capabilities, and is also OLE/DDE compliant. The product is approximately \$200 per person.

3.1.5 Visual Basic/Paradox

PACE and the new Hanford Info are both Hanford-developed Visual Basic applications. PACE uses Paradox, and Hanford Info uses Access as the backend database, to handle the directory management capability. Users view lists of documents via the Visual Basic application, and after selecting the document, the application connects to the server and retrieves the document. By connecting to the server only when necessary, users are prevented from viewing files except through the application. This enhances file control, and provides a means of tracking access and other metrics.

3.1.6 Windows File Manager

The Windows file manager program (winfile.exe) is available to all Windows users. It provides drag-and-drop capabilities. For instance, you can drag a file to a printer icon, to print it. It also provides the capability to launch the file's associated application.

To automatically launch applications, all that is needed is to run "RegEdit" to set up pointers to the proper programs to be associated with the various file extensions (see Appendix A), then use File Manager to *associate* the applications to the desired file extensions.

3.1.7 Lotus Notes

Notes is a "groupware" product. It was specifically designed to be used by groups of users. There is robust security built into the product (perhaps the best in the industry). There is also electronic approval stamping.

- To "electronically approve" a record/document, a user places the user-id stamp on the record. The server's name/address book verifies the user-id key, and confirms its authority. Lotus Notes software ensures that only the user can use his/her associated user-id.

Notes has integrated TOPIC for full text search and retrieval capabilities. Notes also allows responses to documents. Notes includes a forms design system tied to a proprietary database. It comes with several sample applications, including a discussion database. Add-on products include "Lindy", a Notes document scanning/management product, database interfacing (to SQL databases), and work flow (i.e., Action Workflow).

3.1.8 Viewstar

Viewstar is currently in use at Hanford, for the ISEARCH project. Viewstar uses Gupta SQL (and will soon use Windows NT/SQL) to manage indexes to image files. Future plans for Viewstar include integrating OCR capability, text file attachment, and full text searching (possibly using Fulcrum or TOPIC).

3.1.9 Future Trends

- User level security will be implemented at Hanford. This is a capability of the existing network software. It controls who can view, edit, use, and/or delete specific files. Using this security properly, a simple file management process could be established for the site.
- Windows NT will be incorporating a file search capability within its file directory services.
- WordPerfect, recently acquired by Novell, is developing an enterprise-wide document management system, similar to Interleaf. This holds promise because WordPerfect is widely used at Hanford. The office product promises to integrate version control, access control, network management of files, SGML, word indexing, viewing/printing, and a variety of other features.
- Attendance at a February TOPIC user's conference provided insight into future trends for text retrieval. User profiles can be established, to profile individual user's needs and desires for certain types of documents. When those documents become available (via librarian services or real-time news wire feeds), the user is automatically notified and provided with a connection to the appropriate documents.

Other trends include following the Windows NT bandwagon, using client/server technology. A variety of real-time news wire services is becoming available for subscription. TOPIC is also developing means to access multiple TOPIC-created infobases, as well as non-TOPIC created infobases.

3.1.10 Industry News

The following information is documented in this section for reference purposes. Due to resource limitations, there was no follow-up or detailed evaluation performed.

- WordPerfect has developed IntelliTag. WordPerfect also has QuickIndex. IntelliTag is an SGML add-on product, to allow users familiar with WordPerfect to smoothly switch to SGML. Version 2.0 will be a solid SGML product, and is expected to be available in October 1994, as part of WordPerfect 6.*. The QuickIndex product, which is already integrated with WordPerfect 5.2, is capable of indexing documents in multiple drives and subdirectories, for faster word search capability than its sequential search "find" command.
- Adobe Acrobat has imbedded the TOPIC API. A major system for Coca Cola, and Hewlett Packard's New Wave use Saros, but replaced Fulcrum with TOPIC. X-Caliber (based in San Diego) is more of a vertical-oriented product. It has been integrated with the Calera OCR product.
- OpenText is a text retrieval competitor of Verity and Fulcrum. It is a client/server application, and the API is available. It has a very fast search speed. It specializes in SGML text search/retrieval.

It is capable of searching for simple words, phrases (including traditional noise words), boolean, field (title, author, etc.), and proximity searching. Search results can be displayed for all documents, with specific attributes shown (such as title, summary paragraph, etc.) Can view specific fields in results and/or documents (i.e., title and author). Thus, can view result list as words (with surrounding text), documents, etc.

A demonstration was given on February 14, 1994 (coordinated by Kurt Conrad). The company was originally formed to solve Oxford College's challenge to have the capability to make timely revisions to their 70-million word Oxford dictionary. (Previous editions took 30 years to complete).

- The demonstration consisted of 25mb of NY Times articles in tagged text format (similar to SGML). A second demo using SGML was also shown. User interface was UNIX's Motif.

3.3 File Management and Information Repositories

Creation of centralized repositories offers an opportunity to reduce redundant information, and also reduce the need for re-inventing documentation. For instance, all documents could attach a single file, containing acronyms and definitions, instead of creating separate lists for each document. Documents that refer to others can have links to the documents built into the file, and table of contents can link to the associated sections. These capabilities will simplify both the tasks of authoring, and of user access/retrieval.

For example, a link to common procedure could be created. Multiple documents could imbed the common procedure, without the need for retyping. All updates to the common procedure would automatically show in all documents linking to it.

Alternatively, the link would automatically transfer to the appropriate procedure, without the user needing to know how to locate it.

However, access to all information creates new difficulties. Allowing a user to perform massive searches against large repositories could create excessive network traffic. Also, although a specific document may not be "sensitive", an entire collection of documents may be sensitive, because of the "total picture" it may provide. These issues will become increasingly important as Hanford opens up the HLAN to new Hanford contractors, and to the public.

3.4 Networking of Functions

Creating network-accessible compound documents is not simple. The location of each component must be identified (its full network path and file name). Additionally, security issues and version management need to be considered. For example:

- Can a user have access to the compound document, but not to one of its components?
- If a component changes, should the master document be notified and/or be changed?

Other issues include determining the frequency of rebuilding or revising indexes, forcing user check-in of documents, etc.

4.0 PROCESS FOR DEVELOPMENT

4.1 Responsibilities

- Authors should be able to create and save documents.
- Reviewers should have access capabilities to draft documents, with the capability to annotate.
- Approvers should have the capability to view documents, and to approve with comments.
- Publishers should be capable of revising (to add standards); and "publishing" documents into generally-accessible areas. Users should be responsible
- Users should have general access to all active documents, except those that are restricted.

4.2 Funding Identification

Pilot projects need to be identified and funded. Past pilots include ETEXT (aka SHARE), BSEARCH, and PACE.

For site-wide implementation, a general funding source needs to be identified. It is not appropriate to have one organization fund a site-wide system resource. The funding organization would have biased control over the administration, and funding would not necessarily be stable from year to year. There are exceptions to this (Hanford Info, SHARE, Soft Reporting, etc.) but even the exceptions should be revisited to identify or create a "burdened overhead" accounting process. PNL's SARIS system charges a monthly user access fee. Another method may be to charge a flat fee per organization based on the number of people within the organizations.

4.3 Computer Program Interfaces

Interfaces to WordPerfect and cc:Mail are preferred for file management. This will enable viewing, printing, and editing of files. A separate viewer, to prevent editing, is also necessary, to prevent the Hanford Info "flaw" of allowing temporary modification of data prior to printing. WordPerfect markets the "Envoy" software product for viewing WordPerfect (and 10+ other formatted) files.

4.4 On-line Help

The software should be user-intuitive, context sensitive, and should match the Windows "look and feel". User guides with full word search capability are preferred. Common tips, errors, questions, contact points, etc. should be addressed in both the software help, and the Hanford's HUH system.

4.5 Training

On-site training should be provided, to minimize the time and expense of off-site travel.

4.6 Defining Automatic Metrics Collection

To justify additional development and enhancements, statistics are necessary. The statistics should be maintained from the onset:

Metrics should be collected on:

- Total number of documents, total size of documents (bytes and pages)
- Average size of documents (bytes and pages)
- Users (total number, average number, user location/area)
- Access times (by hour, highlighting peak times)
- Access to specific documents, highlighting the most heavily used)
- Document revisions
- System availability and unavailability, errors encountered, etc.
- Customer questions, problems encountered, etc.
- System enhancements, upgrades, bug fixes, etc.
- Other statistics, as needed.

4.7 Technical and User Support

The amount of support necessary should be minimal, to keep the ongoing costs low, and to speed wide acceptance of organization control of file management.

4.8 Administrative Support

The network requirements, training, and end user support for the file management solutions are expected to comprise a minimum of two administrators. One will have primary duties, and the other will serve as a backup. Security for file access, daily administrative duties (review of logs and errors), statistical tracking, training coordination, etc. are some of the tasks that administrators would do.

APPENDIX A: File Naming Standards

Structured naming conventions will simplify the retrieval of files via directory listing methods, and also aid in drag-and-drop launch methods that are available through Windows technology. By reserving file extensions for specific application-created files, users can set up the drag-and-drop associations, simplifying the access to files. The following table shows some of the file extensions, and their associated software applications.

arc	pkzip	nai	Windows WINSTALL program
arj	pkzip	ns2	Lotus Notes
art	scrap book	nsf	Lotus Notes data file
avi	Windows video file (using mplayer)	nsq	Lotus Notes
bas	basic	nsh	Lotus Notes
bat	DOS batch run programs	ntf	Lotus Notes Template file
bif	Binary Initialization File. File extension for WordPerfect 6.* files.	pcx	MS Paintbrush graphics file
bin	DOS programs	pic	Lotus 1-2-3 or Symphony picture
bmp	any bitmapped windows image (i.e., from paint brush)	ppt	Powerpoint file
cal	Windows calendar	prf	WordPerfect Grammatik preference file
cca	a cc:Mail Archive file	prn	any ASCII tabular text
cgm	ANSI Metafile, Freelance	rec	Windows Recorder
cgn	any ASCII structured text	rft	IBM DisplayWrite
ch3	harvard	rip	Lotus Notes error file
clp	Windows clipboard file	rle	a RLE formatted image file
com	DOS programs	rpt	any ASCII tabular text
crd	Windows card file	rtf	any rich text formatted file
dll	Windows dynamic link library	sam	Lotus Ami Pro
doc	Lotus Manuscript, Borland MultiMate, MS Word, WordPerfect 6.*	sh3	Harvard Graphics
dot	MS Word	stf	Lotus Agenda
drw	Micrographics Designer	str	any ASCII structured text
exe	DOS programs	sy3	Harvard Graphics
fon	a Windows font file	syd	Windows Notepad file
fot	a Windows font file	sys	A DOS or Windows system file
frm	Visual Basic	tab	any ASCII tabular text
gif	a GIF image file	tif	any bitmapped scanned TIFF image
gim	Power Point graphics	trm	Windows Terminal
gix	Power Point graphics	ttf	a Windows font file
gmf	ANSI Metafile, Freelance	txt	any ASCII tabular text
gna	Power Point graphics	wav	Windows sound file
gnt	ddGantt	wcm	a WordPerfect 6.* macro
gnx	Power Point graphics	wk1	Lotus 123 or Symphony worksheet
grp	Windows group files	wk3	Lotus 123 or Symphony worksheet
hlp	Windows Help	wks	Lotus 123 or Symphony worksheet
ico	Windows Icon	wp	WordPerfect files
img	an image file	wpd	a WordPerfect 6.* document
ini	Windows ini	wpg	a WordPerfect Graphics file
jpg	a JPEG image file	wpm	a WordPerfect 5.* macro
ltr	any ASCII structured text	wpt	a WordPerfect 6.* template
lzh	an LZH formatted compressed data file	wr1	Lotus 123 or Symphony worksheet
mac	a Macintosh image file	wri	Windows Write
mak	Visual Basic	wrk	Lotus 123 or Symphony worksheet
mmm	a Windows Multi-Media Movie file	wta	a WordPerfect 6.* Text Art file
mpc	MS Project	xla	MS Excel
mpp	MS Project	xlc	MS Excel
mpv	MS Project	xlm	MS Excel
mpw	MS Project	xls	MS Excel
mpx	MS Project	xlt	MS Excel
		xlw	MS Excel
		zip	pkzip formatted compressed data file
		zoo	zoo formatted compressed data file

APPENDIX B: Server Partitioning

The following example has been proposed to BCSR/ISS as a way to create an ISS technical reference library for ISS projects and other ISS documents.

Drive Path	Assigned Responsibility	Purpose	Read-Only Share Name	Read-only Password	Read/Write Share Name	Read/Write Password
C:\ISS Directories may include: \ALL-ISS \ISS-PAGR \Software (sub-directories for various server software, such as Adv-Rev, dd:GANTT, Harvard, Paradox, and PX-Scheduling) \62*** (sub-directories for each org)	JJ Ratcliffe	All-inclusive Common ISS File server area	ISS	alpha	ISS_W	ein
C:\ISS\ALL-ISS Directories may include: \ALT-ANAL \DOCUMENTS \ECCEL \MAIL \MEETINGS \MISC \PX-SCHED \SAFETY \SKILLMTX \STNDARDS \TECH-TIP \TRIP-RPT \WCC		Place for JJ Ratcliffe's and PPC's ISS-wide administrative files, communication, etc.				
C:\ISS\62100 Suggestion: Make separate directories for each project.	RT Hirano	Scientific & Engineering Software Applications				
C:\ISS\62200 Suggestion: Make separate directories for each project.	RL Larson	Information Systems Planning and Development				
C:\ISS\62220 Suggestion: Make separate directories for each project.	DA Guettler	Waste Tracking and Transportation Automation Systems				
C:\ISS\62230 Suggestion: Make separate directories for each project.	DA Guettler	Environmental Systems				
C:\ISS\62250 Suggestion: Make separate directories for each project.	AJ Calapristi	Administrative Systems Development				
C:\ISS\62260 Suggestion: Make separate directories for each project.	PR Edwards	Site Management Information Systems				
C:\ISS\62280 Suggestion: Make separate directories for each project.	SY Bian	LAN Systems Development				
C:\ISS\62300 Suggestion: Make separate directories for each project.	DD Ripplinger	Information Systems Enhancement & Scheduling				
C:\ISS\62310 Suggestion: Make separate directories for each project.	JS Corder	Distributed Systems Support				
C:\ISS\62320 Suggestion: Make separate directories for each project.	TJ Robison	NOMAD2 & MSA Systems Support				

Drive Path	Assigned Responsibility	Purpose	Read-Only Share Name	Read-only Password	Read/Write Share Name	Read/Write Password
C:\ISS\62330 Suggestion: Make separate directories for each project.	RA Richeson	DB2 Systems Support				
C:\ISS\62340 Suggestion: Make separate directories for each project.	WC Winters	LAN Systems Support				
C:\ISS\62350 Suggestion: Make separate directories for each project.	JM Hausel	Facilities Systems Support				
C:\ISS\62400 Suggestion: Make separate directories for each project.	RD VanHouten	Standards & Controls				
C:\ISS\62420 Suggestion: Make separate directories for each project.	ER Carroll	HIMACS Systems Development				
C:\ISS\62600 Suggestion: Make separate directories for each project.	RB Bass	Chemical & Waste Mgmt Software Applications				
C:\ISS\62620 Suggestion: Make separate directories for each project.	MJ Winkelman	Laboratory Information Systems				
C:\ISS\62700 Suggestion: Make separate directories for each project.	FK Mangan	Safeguards & Security Software Applications				
C:\ISS\62800 Suggestion: Make separate directories for each project.	RD VanHouten	System Development Framework				